Ascential Technical Bulletin

Part No. 74-0145

DataStage Pivot

This technical bulletin describes Release 1.0 of the DataStage Pivot stage, formerly called Pivot Table. Pivot is an active stage that pivots an input table by mapping sets of input columns to a single output column.

Printing History

First Edition (74-0145) for Release 1.0, December 1999
Updated for Release 1.0, August 2002
Updated for Release 1.0, August 2003

How to Order Technical Documents

To order copies of documents, contact your local Ascential subsidiary or distributor, or call our main office at (508) 366-3888.

Documentation Team: Marie E. Hedin

Introduction

This technical bulletin describes the following for Release 1.0 of DataStage Pivot for DataStage Release 7.0:

- Functionality
- Configuration requirements
- Installation
- Pivoting data
- Examples

Pivot is an active stage that maps sets of columns in an input table to a single column in an output table. This type of mapping is called pivoting.

This stage pivots horizontal data, that is, columns within a single row into many rows. It repeats a segment of data that is usually key-oriented for each column pivoted so that each output row contains a separate value.

An input column set can consist of one or more columns. The pivoting usually results in an output table that contains fewer columns but more rows than the original input table.

This stage has no stage or link properties. It merely maps input rows to output rows.

For more information about defining input and output columns, see *DataStage Server Job Developer's Guide.*

Functionality

Pivot has the following functionality:

- Supports horizontal pivots.
- NLS (National Language Support). For information, see *DataStage NLS Guide*.

The following functionality is not supported:

- Compatibility with DataStage releases before 7.0.
- Vertical pivots, that is, mapping vertical data in many rows into a single row. (Vertical pivots group one or more columns and map these columns to many columns in the grouped row in an output table.)
- A custom user interface.

Installing the Plug-In

For instructions and information supporting the installation, see *DataStage Plug-In Installation and Configuration Guide*.

To start the plug-in editor from the DataStage Designer, do one of the following:

- Double-click the stage in the Diagram window.
- Select the stage and choose **Edit ➤ Properties…** from the shortcut menu.
- Choose **Edit** from the main **Designer** menu.

Pivoting Data

A horizontal pivot maps columns within a row into many rows, that is, it repeats a segment of data for each column pivoted. The data is usually key-oriented.

Use the **Derivation** field in the output link column grid to specify the pivots. An empty field indicates that there is an input column name with the same name as the output column. This input column is mapped to the corresponding output column. For more information about defining input and output columns, see *DataStage Server Job Developer's Guide.*

Single Derivation. If the **Derivation** field for an output column lists a single column name, the input column having the same name as that specified in the **Derivation** field is mapped to this output column. Any column having a single derivation is treated as a key and is likewise projected to each output row that is derived from the single input row.

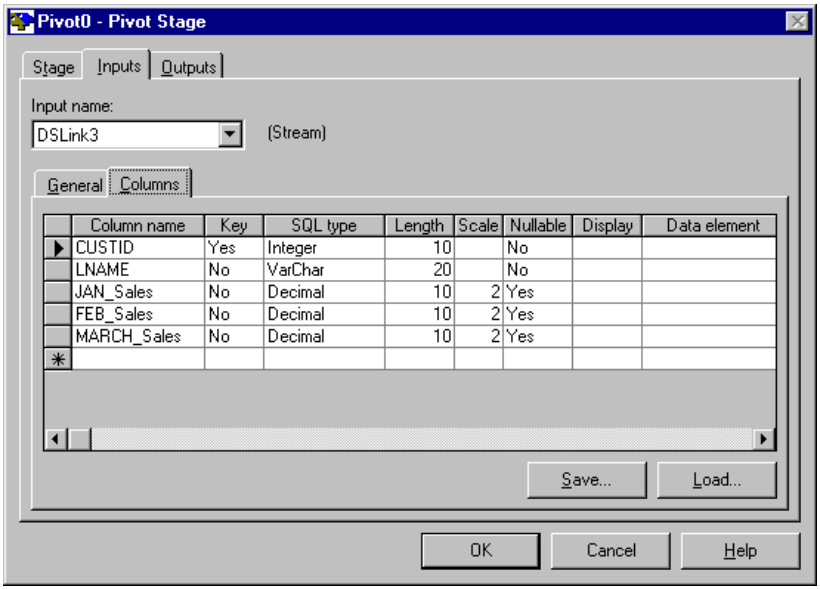
Multiple Derivations. When an output column is derived from more than one input column, that is, more than one input column name is listed in its **Derivation** field, an output table with more rows than the input table results. Each input column specified in the **Derivation** field for the output columns is mapped to the output column. A new row is created for each of the specified input columns.

Examples

The examples described in the following sections show a pivot on the first quarter sales data for a particular enterprise. These examples illustrate the concepts for a horizontal pivot.

Input Link Columns

The input link **Columns** page contains three input columns with sales data: JAN_Sales, FEB_Sales, and MARCH_Sales. The columns are as follows:



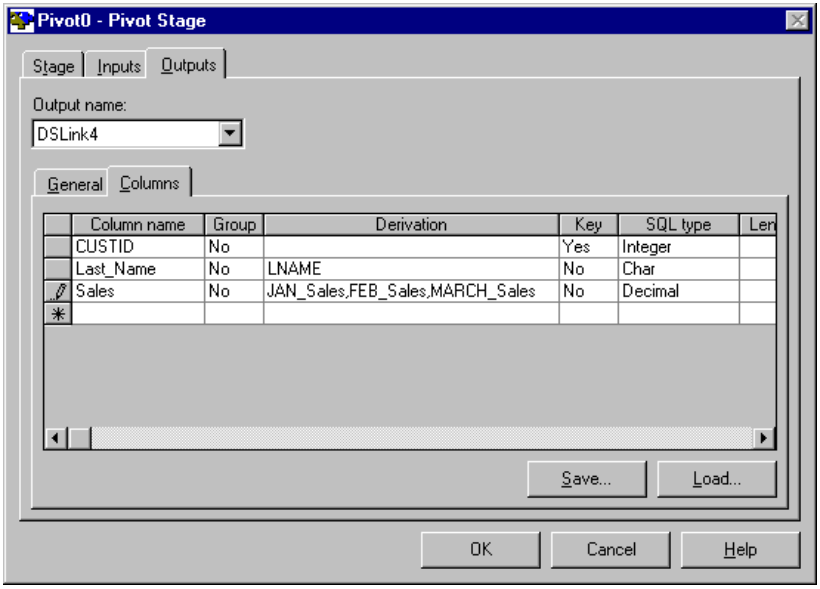
The data for the source rows for the input columns looks like this:

Input Source Rows

CUSTID	LNAME	JAN_Sales	FEB_Sales	MARCH_Sales
100	Smith	$1,234.00	$1,456.00	$1,578.00
101	Yamada	$1,245.00	$1,765.00	$1,934.00

Output Link Columns

The output link **Columns** page contains a Sales column derived from the three input columns: JAN_Sales, FEB_Sales, and MARCH_Sales. The columns are as follows:



The output column that is derived from a single input column is a key value. The key value is repeated in each row that results from the corresponding input row.

The maximum number of output rows that result from a single input row is determined by the output column that is derived from the most input columns. The three output rows of sales data that result from each input row in this example are as follows:

Output Target Rows

CUSTID	Last_Name	Sales
100	Smith	$1,234.00
100	Smith	$1,456.00
100	Smith	$1,578.00
101	Yamada	$1,245.00
101	Yamada	$1,765.00
101	Yamada	$1.934.00

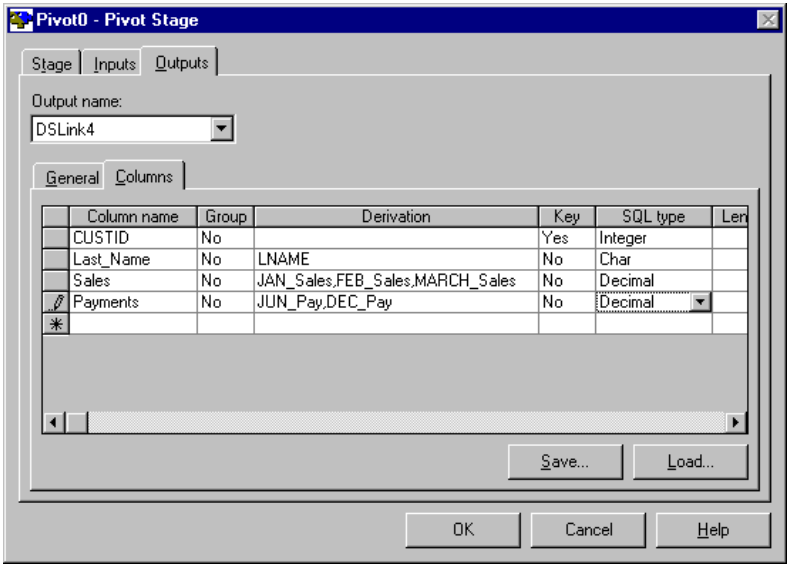
If the pivot includes any derivations with fewer than the maximum number of output rows but more than one row, the output row contains a null value for each column where a derivation is not available.

As an example, let's assume that the customer is required to make payments on his account twice a year, in June and December. The source data might look like this:

Payments Example

CUSTID	LNAME	JAN_Sales	FEB_Sales	MARCH_ Sales	JUN_Pay	DEC_Pay
100	Smith	$1,234.00	$1,456.00	$1,578.00	$6,298.00	$7,050.00
101	Yamada	$1,245.00	$1,765.00	$1,934.00	$7,290.00	$7,975.00

Suppose the output link contains an additional derivation for payments:



The output data in the target rows after the pivot looks like this:

Output Data in Target Rows After Pivot

CUSTID	LNAME	Sales	Payments
100	Smith	$1,234.00	$6,298.00
100	Smith	$1,456.00	$7,050.00
100	Smith	$1,578.00	null
101	Yamada	$1,245.00	$7,290.00
101	Yamada	$1,765.00	$7,975.00
101	Yamada	$1,934.00	null